第42卷，第11期
2022年11月

光 谱 学 与 光 谱 分 析
Spectroscopy and Spectral Analysis

Vol. 42，No. 11，pp3336-3344
November，2022

单类分类方法结合光谱分析在食品真实性鉴别中的应用

唐逸芸[1,4]，刘 芮[2]，王 潞[2]，吕慧英[1,4]，唐忠海[1,4*]，
肖 航[1,3]，郭时印[1,4]，范 伟[1,4*]

1. 湖南农业大学食品科学技术学院，湖南 长沙 410128
2. 云南省烟草公司保山市公司，云南 保山 678000
3. Department of Food Science，University of Massachusetts，Amherst，MA 01003，USA
4. 湖南省菜籽油营养健康与深度开发工程技术研究中心，湖南 长沙 410128

摘 要 近年来，假冒伪劣食品已日益成为广大消费者密切关注的问题，食品真实性评估是缓解这一问题、保护公众健康的有力手段。在仪器设备和样品处理的高要求下，现代检测技术通常需要大量时间和金钱的成本消耗，而如今食品掺假手段不断变换，花样日益翻新，使得这类检测技术存在一定的局限性。为促进食品安全质量监管的效率和水平提高，为监管工作提供有力的科学技术支撑和保障，需要寻求新型检测技术。光谱分析技术，以操作简单、快速无损的优势近年来被广泛应用，作为一种间接分析技术，结合数据统计学中的分类方法建立模型后更能有效进行真假鉴别。在分类方法中，由于现实生活中五花八门的掺假类型以及在真假样本数量差异大的情况下，常用的分类方法效果可能出现偏差。但单类分类方法(one-class classification)是一种只针对一类实例建模分析，以特定的置信水平固定目标样本类的边界，对新样本的类别进行判定的方法，利用这一特点能有效区分不同于真实样本的数据，大大减少了检测的工作量，在食品掺假检测应用领域有一定的发展潜力。对近年来模式识别中的分类方法——单类分类方法进行了综述。通过阐述光谱分析结合分类方法用于食品掺假检测的必要性，比较在同一情形下多类分类方法和单类分类方法的判别率，简介单类分类方法的特点，并重点介绍几种常见的单类分类方法如数据驱动的簇类独立软模式(DD-SIMCA)、单类偏最小二乘(OCPLS)、单类支持向量机(OCSVM)以及单类随机森林(OCRF)，论述单类分类方法在食品真实性鉴别中的应用，具体在食用油，乳制品，饮料，保健品，香辛料及谷物方面进行了阐述，还分析了当前单类分类方法存在的问题，最后对该技术的应用前景进行展望，为食品认证分析提供了一定的理论依据。

关键词 单类分类方法；模式识别；光谱分析；食品掺假

中图分类号：TS207.3 **文献标识码**：R **DOI**：10.3964/j.issn.1000-0593(2022)11-3336-09

引 言

自2008年出现三聚氰胺重大食品安全事故以来，人们对食品真实性问题高度关注，食品欺诈是一种以经济利益为驱动的故意行为，包括“故意替换、添加、篡改或虚报食品成分或食品包装或进行虚假宣传，或关于产品的虚假以及误导性声明”[1]。其中掺假作为欺诈的一种类型，因掺假物的种类性质不同往往会引发食品安全问题，面对全球食品掺假的频繁发生，促使各国政府更加重视食品真伪鉴别，在我国，《食品安全法》中对食品生产过程中的食品掺杂掺假、标签虚假和针对监管的各类信息欺诈的违法行为有详细规定[2]。为了保障食品安全和消费者权益，发展有效的检测方法至关重要。

常用的色谱和质谱等基于化学成分分析的检测方法一般包括复杂的前处理，还存在着检测周期长、需消耗有毒有害

收稿日期：2021-10-12，**修订日期**：2022-02-25

基金项目：国家自然科学基金面上项目(31671858)，湖南省自然科学基金青年项目(2017JJ3107)，湖南省自然科学基金面上项目(2019JJ40114)，湖南省教育厅优秀青年项目(20B286)，湖南省高新技术产业科技创新引领计划项目(2020NK2005)，云南省烟草公司科技项目(2019530000241020，2021530000242040)资助

作者简介：唐逸芸，女，1999年生，湖南农业大学食品科学技术学院硕士研究生 e-mail：eviantyy420@163.com
*通讯作者 e-mail：tangzh@hunau.edu.cn；weifan@hunau.edu.cn

化学试剂、检测成本高、需破坏样本等缺点。随着近代仪器分析的飞速发展，无损快速检测成为食品认证的重要研究方向[3]。光谱技术，如常用的近红外、中红外以及拉曼等振动光谱技术，有弥补传统检测技术的缺陷的可能性，它们凭借着需要较少的制备样品时间，以及快速和、非破坏性和绿色环保等特点，近年来，多被使用为检测样本中掺杂物的替代分析方法[4]。

光谱技术存在特异性低的缺点，如果掺入物的成分几乎接近于原物，两者的光谱差异在肉眼下看起来很难区分，需要借助数学建模将这些信息放大从而找到两者之间的区别，这种复杂的统计学方法是根据样品的相似性将光谱信息(即每个波长的强度)转换成新的变量或类别响应。在过去十年中，已有许多统计学方法与分析测量相结合，被开发应用于质量评估、产品可追溯性，地理来源的定义和检测食品真伪[5]。如主成分分析(PCA)、聚类分析(HCA)等广泛使用的非监督方法是简单有效的分类方法，在没有任何数据先验的情况下通过降低数据维数来识别样本之间的异同[6]。另一方面，线性判别分析(LDA)、偏最小二乘判别分析(PLS-DA)和簇类独立软模式(SIMCA)等监督方法是基于来自特定样本的先验信息生成分类模型，分类用途更广。支持向量机(SVM)、人工神经网络(ANN)、随机森林(RF)等机器学习通过学习如何组合输入信息对从未知数据做出有用的预测。

Oliveri 等[7]得出结论，在正确定义所有的类别，且包含的样本代表每个类的前提下，传统的分类方法使用所有类的贡献，在两个或更多个类之间寻找定界符以区分纯样本和掺有多种已知物的掺假样本。Rodionova 等[8]在一篇关于应用于食品认证的化学计量学方法的详细综述中表示，像判别分析这种分类方法常在代谢组学，基因组学和其他组学中应用，至于认证问题，单类分类方法显示更可靠的结果。

1　单类分类方法

1996 年，Moya 等[9]在研究工作中首创了单类分类这一术语(one-class classification，OCC)，2001 年，Tax[10]进一步阐述和总结了此方法，表明单类分类方法已经成为模式识别的一个重要分支。不同的研究人员根据应用场景的不同来表示类似的概念，如奇异值检测、新奇检测或概念学习等。近年来，单类分类法受到越来越多的关注[11]，其最终结果是回答决策问题中新样本是否属于目标类。关于其分类，根据原理大致可分为四类：第一类是密度估计法，第二类基于神经网络的方法，第三类是基于聚类的方法，第四类是基于支持域的方法[12]。

为了进一步阐述单类分类方法在掺假鉴别中的作用，图 1 模拟了同一情景下不同分类方法的比较情况：

(1) 紫色的圆点表示有一定数量基础的真实样本 T。

(2) 蓝色，绿色，黄色，红色的圆点分别代表不同较少数量的掺假类别(A，B，C，D)，分别将其加入到真实样本 T 中。

(3) 用多类分类方法分类建模后得到不同模型 TA，TB，TC，而未知样本类别 D 未能识别出来。

(4) 用单类分类方法分类建模后得到两大类模型，真实类别 T，其余为掺假类别 ABCD。

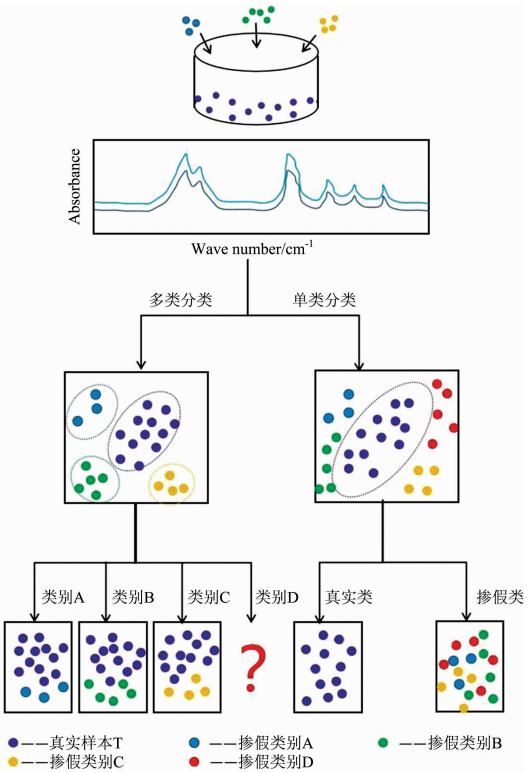

图 1　不同分类方法的示意图

Fig. 1　Schematic diagram of different classification methods

采用多类分类方法进行分类时，在含有已知掺假样本的情况下需要建立不同的模型分别将它们识别出来，但如果掺入未知样本，由于事先并未对其进行过训练，使用该方法效果不佳。相比之下，单类分类方法通常是检测多重掺假的更好选择，因为它只需要用真实的样本来建立分类模型，该模型可以识别任何不同于此的样本为掺假样本，不仅是图中所示的这几种，之后掺入的任何不同于真实样本(E，F，G，…)的都会归类为掺假样本，大大减少了分类工作。

结合高斯函数原理，采用 matlab 分别模拟了三组数量为 1000 的光谱数据，一组为含三个峰的真实样品(图 2)，对于掺假组，为了更好地比较两种方法的分类结果，采用了两种掺假形式，一种是模拟了掺入有三种不同种类的物质(图 3)，而另一种模拟了掺入一类样品(图 4)，取不同数量的真假数据(5/50/500/1 000)，模拟在样品平衡以及差异值很大的情况下，用 PLS-DA 法和 OCPLS 法分别代表多类分类方法和单类分类方法来验证分类效果，结果见表 1 和表 2，其中敏感性表示模型正确分类目标样本的能力，特异性表示模型正确分类非目标样本的能力。

如表 1 和表 2 所示，无论在同类还是不同类的情况下，当真实样本数量极少时，单类分类方法无法识别大量的掺假样本，而多类分类方法效果很好，这也表明单类分类方法要求一定的真实样本数据，而随着真实样本的不断增多，单类分类方法的结果出现逆转，即使在掺假样本极少的情况下，依然有 100%的分类结果，证明了这种方法在处理极端值问题

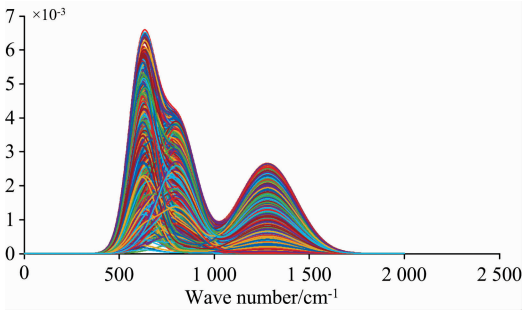

图 2 真实样本模拟数据图

Fig. 2 Simulated data graph ofactual sample

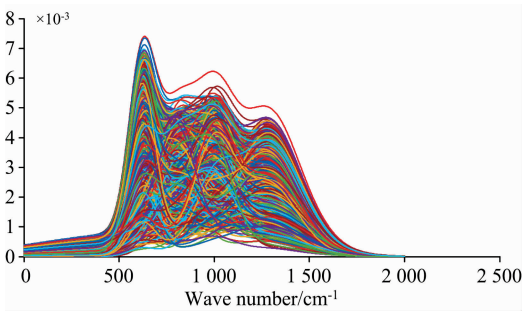

图 3 掺假样本(三种)模拟数据图

Fig. 3 Simulated data graph of adulterated samples (three classes)

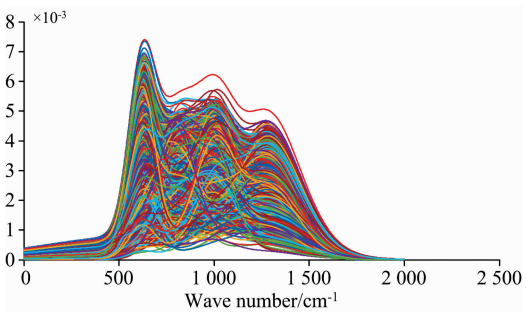

图 4 掺假样本(一种)模拟数据图

Fig. 4 Simulated data graph of adulterated samples (one class)

表 1 模拟不同种类掺假物的分类结果

Table 1 Classification results of simulated adulterants of different classes

真/假	PLS-DA		OCPLS	
	敏感性	特异性	敏感性	特异性
5/1 000	0	100%	100%	0
50/1 000	0	100%	100%	100%
500/1 000	100%	83.4%	73.8%	100%
1 000/1 000	100%	75.6%	83%	100%
1 000/500	100%	70.4%	83%	100%
1 000/50	100%	60%	83%	100%
1 000/5	100%	80%	83%	100%

的优越性。对比表 1 和表 2，在同等数据数量的情况下，掺入有不同种类的样本时，PLS-DA 方法的分类结果低于掺入同类样本的结果，而 OCPLS 不受掺入物类别影响。

表 2 模拟一类掺假物的分类结果

Table 2 Classification results of simulated adulterants of one class

真/假	PLS-DA		OCPLS	
	敏感性	特异性	敏感性	特异性
5/1 000	0	100%	100%	0
50/1 000	0	100%	100%	100%
500/1 000	100%	84.4%	73.8%	100%
1 000/1 000	100%	76.5%	83%	100%
1 000/500	100%	75%	83%	100%
1 000/50	100%	70%	83%	100%
1 000/5	100%	80%	83%	100%

归纳多类分类方法的不足：首先，它依赖于定义明确的类别来训练模型，并且决策边界是根据来自每个类的示例样本决定的，还需要有关掺假物的信息[7]，分类结果旨在将未知数据对象分类为几个预定义类别之一(在最简单的二进制分类情况下为两个)。然而当未知数据对象不属于这些类别时，就会出现问题，而现实生活中的掺假物通常是未知的，当处理未知来源的食品完整性问题时，这种监督方法往往会失败。在样本数量相对平衡的理想情况下，分类面位于分类样本之间，能够较好地区分开来，但在样本数量差异较大的不平衡情况下，如果掺假样本太少，传统分类方法的分类面会明显向少的样本侧偏移，导致分类精度较低，而在现实生活中常出现这种情况，比如机械故障检测，网络入侵检测医学诊断问题中，采集的样本大部分都是正常样本，很少出现异常数据。还有一种可能是异常样本量太大，比如在人脸检测，目标检索和字符检测过程中，虽然异常样本容易获得，但是异常样本的类型太多，几乎不可能全部获得。因此，当样本数量不平衡问题严重时，传统分类方法不能获得很高的分类精度。

单类分类方法已被应用于医学问题[13]，人脸图像识别[14]，故障检测[15]，遥感分类[16]等，如下述几种单类方法。

1.1 数据驱动的簇类独立软模式(data-driven-soft independent modelling of class analogy，DD-SIMCA)

SIMCA(簇类独立软模式)是一种基于主成分分析(PCA)的建模技术，采用 PCA 模型参数和 F 检验构造计算 T^2 ucl 和 Q 统计量作为样本分类的新属性，并计算待测样本到各类主成分空间的欧式距离作为判别类别的依据，是一种常用的分类方法[17]。

DD-SIMCA 是对原 SIMCA 关于构建接受边界方式的修改，作为 PCA 和 SIMCA 的结合，用于开发一个决策规则(阈值)，从所有其他样本中划分出目标类。该方法还提供了一个理论上的可能性计算模型的特点，如Ⅰ型错误 α 和Ⅱ型错误 β[18]。训练数据收集在(I ***X*** J)矩阵 ***X*** 中，I 是样本的数量，J 是变量的数量。计算一般分为两步，首先，DD-SIMCA 将 PCA 应用于 ***X*** 矩阵。主成分(PCs)的个数 A 决定了模型

的复杂度，该参数从根本上影响了分类的质量。A 值越大，$\boldsymbol{X}$ 的大部分变化被 PCA 分解解释。同时，在主成分分析中包含多余的成分可能导致模型不仅考虑了主要的类特征，而且还考虑了不相关的噪声。因此，在选择模型复杂度时，往往采用简约原则。第二步，DD-SIMCA 从训练集中计算每个对象的两个距离，即正交(欧氏)距离(OD)和分数(马氏)距离(SD)。SD 表示样本在分数空间中的位置，OD 表示样本到分数空间的距离。DD-SIMCA 发现了表征这些距离分布的参数的数据驱动估计，因此可以为给定的值开发一个接受区域/决策规则[19]。此外，当可选类可用时，DD-SIMCA 提供了计算Ⅱ型 β 误差并构建相应的扩展接受区域的可能性，保证了从备选类中接受样本的风险不大于 β[18]。分类结果用“灵敏性”(sensitivity)和“特异性”(specificity)来描述，或者用传统的统计术语，如Ⅰ型误差 α 和Ⅱ型误差 β。敏感性表示目标类中正确识别的样本的份额。特异性是可选类对象的一部分，它被正确地标识为该类的数量。根据统计学术语，敏感性可定义为 $100(1-\alpha)\%$，特异性为 $100(1-\beta)\%$[20]。

1.2　单类偏最小二乘法(one-class partial least squares, OCPLS)

虽然各种算法已经发展起来，但最常用的仍然是偏最小二乘法(PLS)，它被公认为化学计量学的基石，其理论和性质得到了广泛的研究[21]。

OCPLS 是一种基于 PLS 的特殊分类算法，被认为是 SIMCA 的替代方法。它是一种在化学计量学中日益受到关注的单类方法，作为一种非线性和鲁棒性算法，它可以减少非线性和异常值污染数据集的影响，在建立 OCPLS 模型时，可以得到两种自然距离度量。一个是基于由主要的 OCPLS 成分从一个样本到类的中心跨越的分数距离(SD)；另一个是响应变量为 1 的绝对中心残差(ACR)。可以计算出样本的 Hotelling's T^2 统计量，并通过计算 F 分布可以得到 SD 的置信上限(UCL)，再计算模型残差。样本的残差可以被假定为一个正态分布，有一个估计均值和一个估计标准差。样本的残差可以集中为零均值。因此，可以得到类内样本的 ACR 的 UCL。ACR 值实际上是 OCPLS 回归系数向量上投影的度量，而 OCPLS 模型的一个组成部分可以看作是一个样本在训练集的平均光谱上的投影。一般来说，类内样本的投影到平均频谱上有相当长的长度和分布紧密，这意味着 OCPLS 组件考虑了两者解释方差和预测的紧致性。对于一个样本，过大的 SD 或 ACR 值表明它明显偏离类的大部分。根据 ACR 和 SD 的值，一个未知的测试样本可以分配给一个四组：正常样本(低 SD 和低 ACR 值)，坏杠杆样本(高 SD 和高 ACR 值)，良好的杠杆样本(高 SD 和低 ACR 值)和响应异常值(低 SD 和高 ACR 值)，正常样本被视为真实样本，而其他三种被认为是掺假的或者伪造的样本[22]。

1.3　单类支持向量机(one-class support vector machine, OCSVM)

SVM(支持向量机)的目标是通过最大化分离超平面和数据之间的距离或余量，找到一个泛化误差最小的最优超平面[23]。

单类支持向量机(OCSVM)是原始 SVM 算法的变体[24]。在高维空间中给定一组训练数据，OCSVM 就是在一个变换的空间中寻找超平面，该空间将大部分数据集中的区域与其他地方分开[25]。对超平面参数进行估计，使其与训练数据相关的余量最大化。因此，它是寻找训练点与原点最大间隔或确定包含同一类训练数据的最小超球体的最有效方法之一。数据分类包括检查测试样本是否属于超球体。OCSVM 的运行时性能取决于支持向量机的数量(SVs)，这可能比训练样本的数量少得多。在掺假检测的背景下，从真实样品中识别所有可能的掺假样品是至关重要的。实际上，OCSVM 计算出容纳大多数训练点的“边界”，如果测试样品落在此界限内，则归类为真实样本；否则，它被视为掺假样本。

OCSVM 算法是一种用途广泛的分类器，能够应用于负类样本难以收集的领域中[26]，已用于许多不同的领域，例如工程[27]、地质[28]，提供了有意义的结果。尽管 OCSVM 在解决复杂问题方面表现出色，但在化学中并未得到充分利用[29]，只有少数研究将该算法应用于分析目的，比如使用气相色谱法结合 OCSVM 检测芝麻油中的掺假[30]，还有应用 OCSVM 作为异常值检测器来追踪茶叶的地理来源[31]。目前结合 OCSVM 和光谱学用于食品认证的化学研究还少被涉及，这也是未来趋势。

1.4　单类随机森林(one-class random forests, OCRF)

随机森林(RF)作为通用的集成技术之一，使用随机化产生不同的基于个体树的分类器池[32]。随机森林算法使用了两个强大的随机化过程：打包(bagging)和随机特征选择(random feature selection, RFS)。第一个原则，bagging 是在训练集的引导副本上训练每个单独的树，通常用于在各个分类器之间创建预期的多样性，并且对于不稳定的分类器特别有效，例如基于树的分类器，其中训练集的小变化导致预测的大变化。第二个原则，RFS 是一个随机原则，专门用于树归纳算法。当生长成树时，包括在树的每个节点随机选择特征子集，从中选择分裂测试。RFS 有助于降维，并已被证明比单独 bagging 显著提高随机森林精度。

而单类随机森林(OCRF)，是在随机森林算法的基础上增加了原始异常值生成过程，该过程利用随机森林算法提供的集成学习机制来减少要生成的人工异常值的数量以及生成这些异常值的特征空间的大小[33]。OCRF 方法具有以下优点：(1)组合弱分类器和不稳定分类器的不同集合，明确提高了单个分类器的泛化性能，(2)依据训练样本和特征对训练数据集进行子采样，以便通过控制它们的位置和数量有效地生成输出。

2　单类分类方法在食品掺假中的应用

2.1　食用油

作为人体必需的三大营养素来源之一的食用油，其掺假是消费者和油脂加工业的首要担忧，掺假主要有两种类型，一种是冷榨油和精炼油的混合，一种是用便宜的食物机制代替昂贵的食物机制。

Rodriguez 等[34]用傅里叶变换红外光谱结合 OCPLS 和 SIMCA 检测以 1%，2%，5%和 10%四种不同的比例掺入到芝麻油中四种可能的掺杂物的存在。结果表明，在预测误差

1%～5%内，用这种方法检测掺假的奇亚籽油和芝麻油是成功的，且 OCPLS 比 SIMCA 的鉴别性能稍高一点，也体现了单类分类方法的优越性。Hu 等[35]基于 OCPLS 近红外光谱和荧光光谱数据融合，采用快速分析证实了中国油茶中掺入廉价植物油的可行性。结果表明，鲁棒的 OCPLS 可以检测掺有 2%及以上的包括菜籽油、葵花籽油、玉米油和花生油的廉价油。Neves 等[36]评估了傅里叶变换衰减全反射红外光谱法结合 DD-SIMCA 检测初榨椰子油的掺假性能，通过测定纯油样品和掺有油菜籽油、玉米油、向日葵油和大豆油样品的红外光谱，用单类分类模型来判定初榨椰子油的真实性和掺假性，最后以 88%～100%的灵敏度和 96%～100%的特异性识别掺假油。

然而，单类分类模型也会出现分类效果弱于多分类情况，例如 Gagneten 等[37]于检测菜籽油中的掺假物比较 SIMCA，PLS-DA，DD-SIMCA 和 OCPLS 四种方法，结果表明用 SIMCA，PLS-DA 的准确率略高于 DD-SIMCA 和 OCPLS。分析认为某些波长不包含必要的信息，可能会干扰模型的建立步骤，而如果通过适当选择输入变量，选择与目标样品特性高度相关的波长范围，分类结果可能会得到改善。Yuan 等[38]采用近红外光谱和 OCPLS 法对亚麻籽油进行多重掺假的有针对性的检测，并且设计了一种变量选择方法，以显著减少变量数量，提高掺杂物检测的准确性。

2.2 乳制品

为了达到质量要求标准，乳制品的掺假主要是通过添加化合物完成，一般有为了减少微生物的数量而添加过氧化氢、甲醛或次氯酸钠等被归类为防腐剂的物质，以及添加氯化钠、淀粉或蔗糖等被归类为增稠剂的物质。

Gondim 等[39]提出了一种采用中红外光谱技术和单分类方法对牛奶中掺假成分进行序列检测的方法。模型采用低目标掺假水平，包括甲醛、过氧化氢、碳酸氢盐、碳酸酯以及蔗糖等，因减少了所需要时间及成本和错误的样本数量，这种方法被认为是一种有效的筛选方法。Muller-Maatsch 等[40]将紫外可见荧光和近红外光谱技术与单类分类方法相结合，以区分真正的脱脂奶粉和掺假奶粉，最后有 86%的掺假样品被正确地归类为“不合格”。

2.3 饮品

饮料的掺假主要有两种类型，一种是使用较便宜的水果来代替单一果浆中的主要成分，还有一种是对饮料进行有关成分，真实性或地理起源的错误标签标识。

Xu 等[41]采用傅里叶变换近红外光谱对正宗板蓝根茶的成分和类别模型进行了表征，并对可能的外源性掺假物进行了检测。采用标准正态变换(SNV)得到最精确的 OCPLS 模型。结果表明，SNV-OCPLS 可以检测到板蓝根中掺假量在 5% (W/W)以上的苹果干皮，为板蓝根茶的快速质量控制提供了一种有用的替代工具。Xu 等[42]采用荧光法和化学计量学方法研究了猕猴桃汁中多种廉价物质同时检测的可行性，最后得到了灵敏度为 0.929 的 OCPLS 模型。该方法可以检测出 2%以上的糖浆和人造果粉掺假，为非靶向分析掺假猕猴桃汁提供了一种快速和高灵敏的方法。Miaw 等[43]采用低场核磁共振光谱评价了苹果汁、腰果汁和混合果汁对葡萄蜜酒的掺假。采用 OCPLS，DD-SIMCA 和 PLS-DA 等分类方法进行比较。结果表明，所有单类分类方法均具有良好的性能，分辨率高于 93%，而多类方法分类结果不太满意，这也凸显了单类分类的优势。

2.4 保健品

药用保健食品由于经济效益的原因，经常被添加一些外观相似的廉价材料，如粉末或提取物制造假冒伪劣的药材制剂。

Li 等[44]采用近红外光谱和 OCPLS 建立了来自不同产地的代表性中草药天麻样品的类模型，对芋头淀粉、甘薯淀粉、马铃薯淀粉和黄精粉 4 种常见外源性掺假物进行了非靶向检测。结果表明，经过二阶导数处理后的光谱 OCPLS 模型可以检测出 1.0% 及以上的 4 种掺假物，灵敏度为 0.910 7。Rodionova 等[45]以牛至药材掺假为例，采用判别分析和单类分类法分析了非目标分析在食品欺诈检测中的应用所涉及的化学计量学问题。结果表明，判别方法只是部分适用于解决认证问题，DD-SIMCA 是用于非目标分析的功能强大的分类器。在中国，可食燕窝作为一种珍贵的功能性产品，需要建立一种可靠的方法来快速鉴定。Guo 等[46]采用傅里叶变换红外光谱结合 PCA，LDA，SVM 和 OCPLS 等化学计量学方法，验证了该系统识别的可行性。结果表明，OCPLS 模型的预测灵敏度为 0.937，特异度为 0.886，对商业可食燕窝样品的检测有了进一步的推进。

2.5 香辛料

香料用来给食物调味和改善菜肴的味道，色泽是香料的主要品质属性之一。常见的香料掺假是添加非法染料，人为地提高和保持香料的天然色泽，或掩盖与低价值产品原料的混合。此外，香料的价格通常是由它们的重量或体积决定的，而另一种常见的香料掺假是添加便宜的膨化剂。

Horn 等[47]采用傅里叶变换中红外光谱和 DD-SIMCA，基于不同预处理方法比较，建立了一种辣椒粉掺假的非靶向检测方法，测试含 1% (W/W)苏丹Ⅰ、1% (W/W)苏丹Ⅳ、3% (W/W)铬酸铅、3% (W/W)氧化铅、5% (W/W)二氧化硅，10%的聚氯乙烯，10%的阿拉伯胶的掺假物。随后他们[48]又将核磁共振波谱与单类分类法相结合用于辣椒粉掺假的非靶向检测，建立的单类分类模型灵敏度为 92%，适合掺假筛查和与异常值诊断相结合。

2.6 谷物

谷物的品质由掺入便宜的粉末及添加剂来改变蛋白质含量、淀粉含量或硬度。

Cardoso 等[49]用拉曼光谱结合 OCSVM 和 SIMCA 对木薯淀粉样品进行改性，将掺假物如小麦粉、碳酸氢钠等以 0.5%～50%的范围掺入木薯淀粉中。对这两种化学计量模型进行统计比较，发现 OCSVM 优于 SIMCA，OCSVM 检测掺假率超过 2%的可能性，而 SIMCA 检测掺假率只有 5%。Faqeerzada 等[50]用高光谱短波红外图像结合 DD-SIMCA 对掺入不同比例花生粉的杏仁粉进行了研究，建立了 PLSR 模型来预测杏仁粉中掺假比例。DD-SIMCA 的分类结果对不同的掺假样本验证集具有 100%的敏感性和 89%～100%的特异性。PLSR 分析结果表明，每一种掺杂的杏仁粉具有较高

的判定系数和较低的误差值。Rodionova 等[51]通过对大豆粕进行近红外光谱测量和 DD-SIMCA 进行数据处理，鉴别出三聚氰胺、氰尿酸和混合掺假物，证明了此方法的可靠性。

表 3 整理了上述单分类方法结合光谱分析在食品掺假检测方面应用的相关文献。

表 3　单类分类方法结合光谱分析在食品掺假检测方面的应用

Table 3　Application of one-class classification combined with spectral analysis in food adulteration detection

类别	检测对象	应用技术	分析方法	掺假物质	检测结果	参考文献
食用油	奇亚籽油和芝麻油	傅里叶变换红外光谱	OCPLS，SIMCA	玉米油、花生油、大豆油和葵花籽油	正确识别率都有 94%以上	[34]
	茶油	近红外光谱和荧光光谱	OCPLS	菜籽油、葵花籽油、玉米油和花生油	灵敏度为 95.4%，特异性为 91%	[35]
	初榨椰子油	傅里叶变换衰减全反射红外光谱	DD-SIMCA	菜籽油、玉米油、葵花籽油和大豆油	88%～100%的灵敏度，96%～100%的特异性	[36]
	菜籽油	傅里叶变换红外光谱	SIMCA，PLS-DA，DD-SIMCA 和 OCPLS	玉米油、花生油、大豆油和葵花籽油	SIMCA，PLS-DA，的分类效果高于 DD-SIMCA 和 OCPLS	[37]
	亚麻籽油	近红外光谱	OCPLS	菜籽油、玉米油、葵花籽油，棉籽油和大豆油	正确识别率 95.8%	[38]
乳制品	牛奶	中红外光谱	SIMCA	甲醛、过氧化氢、碳酸氢盐、碳酸酯和蔗糖	82%的正确分类、17%的不确定分类和 1%分类错误	[39]
	脱脂奶粉	紫外-可见、荧光和近红外光谱	SIMCA 和 OCSVM	氯化铵、硝酸铵、三聚氰胺和尿素	总体准确率为 86%	[40]
饮品	正宗板蓝根茶	傅里叶变换近红外光谱	OCPLS	干苹果皮	准确率为 93.6%	[41]
	猕猴桃汁	荧光光谱	OCPLS	糖浆和人造果粉	灵敏度为 92.9%	[42]
	葡萄蜜酒	低场核磁共振光谱	OCPLS，DD-SIMCA 和 PLS-DA	苹果汁、腰果汁和混合果汁	分辨率高于 93%	[43]
保健品	中草药天麻	近红外光谱	OCPLS	芋头淀粉、甘薯淀粉、马铃薯淀粉和黄精粉	灵敏度为 91.07%	[44]
	牛至药材	近红外光谱	PLS-DA，DD-SIMCA	榛子、橄榄叶和迷迭香等	单类分类器功能强大	[45]
	燕窝	傅里叶变换红外光谱	LDA，SVM 和 OCPLS	银耳、琼脂、炸猪皮和蛋清	灵敏度为 93.7%，特异度为 88.6%	[46]
香辛料	辣椒粉	傅里叶变换中红外光谱	DD-SIMCA	苏丹Ⅰ、苏丹Ⅳ、铬酸铅、氧化铅、二氧化硅、聚氯乙烯和阿拉伯胶	所有掺假物的特异性＞80%	[47]
	辣椒粉	核磁共振光谱	DD-SIMCA	偶氮红、甜菜根和漆树粉	灵敏度为 92%	[48]
谷物	木薯淀粉	拉曼光谱	OCSVM 和 SIMCA	小麦粉和碳酸氢钠	可检测掺假率超过 2%的可能性	[49]
	杏仁粉	高光谱短波红外图像	DD-SIMCA	花生粉	100%的敏感性和 89%～100%的特异性	[50]
	大豆粕	近红外光谱	DD-SIMCA	三聚氰胺、氰尿酸和混合掺假物	灵敏度为 98%	[51]

3　结　论

光谱检测技术是现阶段比较常用的检测技术，将其应用于食品质量安全检测中，不仅可以保证饮食安全，还能促进光谱技术的发展。随着多元统计学的不断发展，分类方法在食品质量安全检测方面有了更深更广的发展空间。多类分类方法的分析过程需要复杂的统计方法，精准的建模和完善的算法，而单类分类方法只需要对目标类进行分类，确定好边界后，其余可能不同的样本都将与其分开，大大减少了分类的工作量。在现实生活中，可以先用此方法筛选出掺假的样品，再对掺假样品进行定量调查。

在过去的几年里，新的单类分类算法出现了，并在一些应用领域得到了开发。尽管单类分类领域正在变得成熟，但

仍有几个基本问题有待研究，首先注意的是单类分类的任务是在正常类周围定义一个分类边界，这样它可以从正常类中接受尽可能多的对象，同时最大限度地减少接受异常对象的机会。由于只能确定边界的一边，因此很难根据一个类别来确定边界在数据周围的每个方向上的紧密程度，也更难确定应该使用哪些属性来寻求正常和异常对象的最佳分离。特别是，当数据的边界长且不凸时，所需的训练对象的数量可能会非常高。所以，相对于传统的多类分类算法，单类分类算法将需要更多的训练数据。其次，分类器集成方法需要进一步探索，基于随机子空间新技术值得关注，随机预言集成在多类分类问题上表现得更好，OCRF 是新出现的这一方面的方法，当然，还可以进行新的集成方法研究。而且在 OCS-VM 中使用的内核多数是线性，多项式以及高斯的，研究人员可以专注于有效调整和优化核函数研究一些更具创新性的核形式。开发用于流式数据分析和在线分类的单类分类方法也是值得期待的。总之，没有一篇文献指出单类分类方法要优于多类分类方法，选择最佳的分析和统计方法并不是一件容易的事情，这将取决于具体的食物真实性问题，因为所有的方法都有优点和缺点。目前要做的，是需要不断完善各类单类算法，并与多类分类方法相结合比较，得到对于不同类型样品最适合的算法，取得最优结果，进一步监测食品的质量安全。

References

[1] Spink J, Moyer D C. Journal of Food Science, 2011, 76(9): R157.
[2] SUN Ying(孙 颖). Research on China Market Regulation(中国市场监管研究), 2017, (11): 19.
[3] Danezis G P, Tsagkaris A S, F Camin, et al. TrAC Trends in Analytical Chemistry, 2016, 85: 123.
[4] WANG Xiao-yan, WANG Xi-chang, LIU Yuan, et al(王小燕, 王锡昌, 刘 源, 等). Food Science(食品科学), 2011, 32(1): 265.
[5] Granato D, Putnik P, Kovacevic D B, et al. Comprehensive Reviews in Food Science and Food Safety, 2018, 17(3): 663.
[6] Gomez-Caravaca A M, Maggio R M, Cerretani L. Analytica Chimica Acta, 2016, 913: 1.
[7] Oliveri P. Analytica Chimica Acta, 2017, 982: 9.
[8] Rodionova O Y, Titova A V, Pomerantsev A L. Trac-Trends in Analytical Chemistry, 2016, 78: 17.
[9] Moya M M, Hush D R. Neural Networks, 1996, 9(3): 463.
[10] Tax D M J, Duin R P W. Journal of Machine Learning Research, 2001, 2(12): 155.
[11] Rodionova O Y, Oliveri P, Pomerantsev A L. Chemometrics and Intelligent Laboratory Systems, 2016, 159: 89.
[12] PAN Zhi-song, CHEN Bin, MIAO Zhi-min, et al(潘志松, 陈 斌, 缪志敏, 等). Acta Electronica Sinica(电子学报), 2009, 37(11): 2496.
[13] Irigoien I, Sierra B, Arenas C. Scientific World Journal, 2014, 2014: 730712.
[14] Arashloo S R, Kittler J, Christmas W. IEEE Access, 2017, 5: 13868.
[15] Xiao Y, Wang H, Xu W, et al. Chemometrics and Intelligent Laboratory Systems, 2016, 151: 15.
[16] Li W, Guo Q. International Journal of Remote Sensing, 2010, 31(8): 2227.
[17] Wold S, Sjölröm M. SIMCA: a Method for Analyzing Chemical Date in Terms of Similarity and Analogy, 1997.
[18] Pomerantsev A L, Rodionova O Y. Journal of Chemometrics, 2014, 28(6): 518.
[19] Pomerantsev A L, Rodionova O Y. Journal of Chemometrics, 2014, 28(5): 429.
[20] Rodionova O Y, Balyklova K S, Titova A V, et al. Journal of Pharmaceutical and Biomedical Analysis, 2014, 98C: 186.
[21] Wold S, Trygg J, Berglund A, et al. Chemometrics & Intelligent Laboratory Systems, 2001, 58(2): 131.
[22] Xu L, Yan S M, Cai C B, et al. Chemometrics and Intelligent Laboratory Systems, 2013, 126: 1.
[23] Seo K K. Expert Systems with Applications, 2007, 33(2): 491.
[24] Mahadevan S, Shah S L. Journal of Process Control, 2009, 19(10): 1627.
[25] Kaneko H, Funatsu K. AICHE Journal, 2013, 59(6): 2046.
[26] YIN Chuan-huan, MU Shao-min, TIAN Sheng-feng, et al(尹传环, 牟少敏, 田盛丰, 等). Computer Engineering and Applications(计算机工程与应用), 2012, 48(12): 1.
[27] Zhao Y P, Huang G, Hu Q K, et al. Engineering Applications of Artificial Intelligence, 2020, 94: 103796.
[28] Xiong Y, Zuo R. Computers & Geosciences, 2020, 140: 104484.
[29] Lang R, Lu R, Zhao C, et al. Applied Mathematics and Computation, 2020, 364: 124487.
[30] Zhang L, Huang X, Li P, et al. Chemometrics and Intelligent Laboratory Systems, 2017, 161: 147.
[31] Hong X Z, Fu X S, Wang Z L, et al. Journal of Analytical Methods in Chemistry, 2019, 219: 1537568.
[32] Biau G, Scornet E. Test, 2016, 25(2): 197.
[33] Desir C, Bernard S, Petitjean C, et al. Pattern Recognition, 2013, 46(12): 3490.
[34] Rodriguez S D, Gagneten M, Farroni A E, et al. Food Control, 2019, 105: 78.
[35] Hu O, Chen J, Gao P, et al. Journal of the Science of Food and Agriculture, 2019, 99(5): 2285.
[36] Neves M D G, Poppi R J. Talanta, 2020, 219: 121338.
[37] Gagneten M, Buera M D P, Rodriguez S D. International Journal of Food Science and Technology, 2021, 56(6): 2596.

[38] Yuan Z, Zhang L, Wang D, et al. LWT-Food Science and Technology, 2020, 125: 109247.
[39] Gondim C D S, Junqueira R G, Carvalho De Souza S V, et al. Food Chemistry, 2017, 230: 68.
[40] Mueller-Maatsch J, Alewijn M, Wijtten M, et al. Food Control, 2021, 121: 107744.
[41] Xu L, Fu X S, Fu H Y, et al. Journal of Food Quality, 2015, 38(6): 450.
[42] Xu L, Shi Q, Lu D, et al. Microchemical Journal, 2020, 157: 105105.
[43] Miaw C S W, Santos P M, Silva A R C S, et al. Food Analytical Methods, 2020, 13(1): 108.
[44] Li G F, Yin Q B, Zhang L, et al. Analytical Methods, 2017, 9(12): 1897.
[45] Rodionova O Y, Pomerantsev A L. Food Chemistry, 2020, 317: 126448.
[46] Guo L, Wu Y, Liu M, et al. Journal of the Science of Food and Agriculture, 2018, 98(8): 3057.
[47] Horn B, Esslinger S, Pfister M, et al. Food Chemistry, 2018, 257: 112.
[48] Horn B, Esslinger S, Fauhl-Hassek C, et al. Food Control, 2021, 128.
[49] Cardoso V G K, Poppi R J. Food Control, 2021, 125: 107917.
[50] Faqeerzada M A, Lohumi S, Kim G, et al. Sensors, 2020, 20(20): 5855.
[51] Rodionova O Y, Pierna J a F, Baeten V, et al. Food Control, 2021, 119: 107459.

Application of One-Class Classification Combined With Spectral Analysis in Food Authenticity Identification

TANG Yi-yun[1,4], LIU Rui[2], WANG Lu[2], LÜ Hui-ying[1,4], TANG Zhong-hai[1,4*], XIAO Hang[1,3], GUO Shi-yin[1,4], FAN Wei[1,4*]

1. College of Food Science and Technology, Hunan Agricultural University, Changsha 410128, China
2. Baoshan Tobacco Company of Yunnan Province, Baoshan 678000, China
3. Department of Food Science, University of Massachusetts, Amherst, MA 01003, USA
4. Hunan Engineering Technology Research Center for Rapeseed Oil Nutrition Health and Deep Development, Changsha 410128, China

Abstract In recent years, counterfeit and substandard food products have become an increasing concern to consumers, and food authenticity assessment is a powerful tool to address this problem and protect public health. Under the high requirements of equipment and sample processing, modern detection technologies usually require a lot of time and money cost consumption. However, as food adulteration methods change and become more sophisticated, traditional modern food quality detection technologies have certain limitations. Therefore, it is necessary to seek new detection technology to effectively promote the efficiency and improvement of food safety quality control and provide strong scientific and technological support and protection for regulatory work-spectroscopic analysis technology, which has been used extensively in recent years for its simplicity and rapidity. As an indirect analysis technique, it needs to be combined with classification methods in data statistics to establish models and achieve rapid analysis requirements. Commonly used classification methods are ineffective in the face of the enormous variety of adulteration types in real life and the considerable variation in the number of true and false samples. One-class classification is a method that models and analyses only one class of instances, fixing the boundaries of the target sample class at a specific confidence level for classification and then using the edges of the target sample to predict the class of the new sample, distinguishing it from all other possible objects. Using this feature to effectively differentiate between samples that are different from the actual data, significantly reducing the detection effort, and has some potential for development in food adulteration detection applications. This paper reviewed the one-class classification method, which has been used in pattern recognition in recent years and described the need for spectral analysis combined with classification methods for food adulteration. The classification results of traditional -and one-class classification methods were compared in the same scenario, and the latter's characteristics were briefly introduced. Then, several common one-class classification methods were highlighted, such as data-driven class comparison soft independent modelling (DD-SIMCA), one-class partial least squares (OCPLS), and one-class support vector machines (OCSVM), and one-class random forests (OCRF). The applications of one-class classification methods in the food authenticity identification were also discussed, specifically in edible oils, dairy products, beverages, herbs, spices, and agricultural products. At last, the problems of the current one-class classification were analyzed, and the prospects for applying the technique were outlined. This paper is expected to provide some theoretical basis for food certification analysis.

Keywords One-class classification method; Pattern recognition; Spectral analysis; Food adulteration

* Corresponding authors (Received Oct. 12, 2021; accepted Feb. 25, 2022)

《光谱学与光谱分析》投稿简则

《光谱学与光谱分析》是由中国科协主管,中国光学学会主办,钢铁研究总院、中国科学院物理研究所、北京大学、清华大学共同承办的专业学术期刊。国内外公开发行,从 2004 年起为月刊,大 16 开本,每期 332 页。《光谱学与光谱分析》主要报道我国光谱学与光谱分析领域内具有创新性科研成果,及时反映国内外光谱学与光谱分析的进展和动态;发现并培育人才;推动和促进光谱学与光谱分析的发展。为科教兴国服务。读者对象为从事光谱学与光谱分析的科研人员、教学人员、分析测试人员和科研管理干部。

栏目设置和要求

1. 研究报告 要求具有创新性的研究成果,一般文章以 8000 字(包括图表、参考文献、作者姓名、单位和中文、英文摘要,下同)为宜。

2. 研究简报 要求在前人研究的基础上有重大改进或阶段性研究成果,一般不超过 5000 字。

3. 评述与进展 要求评述国内外本专业的发展前沿和进展动态,一般不超过 10000 字。

4. 新仪器装置 要求介绍新型光谱仪器的研制、开发、使用性能和应用,一般不超过 5000 字。

5. 来稿摘登 要求测试手段及方法有改进并有应用交流价值,一般以 3000~4000 字为宜。

稿件要求

1. 投稿者请经本刊编委(或历届编委)一人或本专业知名专家推荐,并附单位保密审查意见及作者署名顺序,主要作者介绍。文章有重大经济效益或有创新者,请说明,同时注明受国家级基金或国家自然科学基金资助情况。

2. 来稿要观点明确、数据真实可靠、层次分明、言简意明、重点突出。来稿必须是网上在线投稿(含各种符号和外文字母大写、小写、正体、斜体;希腊字母、拉丁字母;上角、下角标位置应标清楚)。中文摘要以 800 字为宜,英文摘要(建议经专业英语翻译机构润色)与中文摘要要对照;另附关键词。要求来稿应达到"齐、清、定",中文、英文文字通顺,方可接受送审。

3. 为了进一步统一和完善投稿方式、缩短论文发表周期,本刊**只接收网上在线投稿**,不接收以邮寄方式或 e-mail 方式的投稿,严禁"一稿多投",对侵权、抄袭、剽窃等学术不端行为,一经发现,取消三年投稿资格。

4. 文中插图要求完整,图中坐标、线条、单位、符号、图注等应标注准确、完整。如作者特殊要求需出彩色插图者,必须在投稿时事先加以说明,并承担另加的彩印费用。图幅大小:单栏图 7.5cm(宽)×6cm(高);双栏图:14cm(宽)×6cm(高);图中数字、图题、表题全部用中文、英文对照,图中数字、中文、英文全用 6 号字。电子文档中除实物图外,曲线图要用 Matlab,Excel,Visio 或 Origin 等软件制作,稿件中图片的原图并转成相应(可编辑)的文件格式(.fig,.xls,.vsd,.opj),非".jpg"格式的文档,随电子版修改稿一同发送到本刊的修改稿专用邮箱。

5. 文中出现的单位必须按"中华人民共和国计量标准"及有关 GB 标准规定缮写。物理量符号一律用斜体,单位符号和词头用正体字母。

6. 名词术语,请参照全国科学技术名词规定缮写。

7. 参考文献,采用顺序编码制,只列主要文献;以 15~20 条为宜。**内部资料、私人通讯、未经公开发表的一律不能引用。**日文、俄文等非英文文献,请用英文表述;中文文献和中文图书采用中、英文对照表述,文献缮写格式请参照本刊。

8. 请在投稿第一页左下角写明投稿联系人的**电话和两个 e-mail,以便及时联系。**

稿件处理

1. 自收到稿件之日起,一个月内作者会收到编辑部的稿件处理意见。请根据录用通知中所提出的要求认真修改,希望修改稿在 30 天内寄回编辑部,并作为作者最终定稿(**当作者接到校样时,以此修改稿为准进行校对,请勿再做大的改动**),若二个月内编辑部没收到修改稿,**将视为自行撤稿处理。**

2. 有重大创新并有基金资助者可优先发表;不录用的稿件,编辑部将尽快通知作者,底稿一律不退,请自留底稿。

3. 来稿一经发表将酌致稿酬并送样刊 2 册。

4. 遵照《中华人民共和国著作权法》,投稿作者须明确表示,该文版权(含各种媒体的版权)授权给《光谱学与光谱分析》期刊社。国内外各大文献检索系统摘录本刊刊出的论文;凡不同意被检索刊物无稿酬摘引者,请在投稿时事先声明,否则,本刊一律认为已获作者授权认可。

5. 修改稿请寄:100081 北京市海淀区学院南路 76 号(南院南门),《光谱学与光谱分析》期刊社(收)

电话:010-62182998 或 62181070 传真:010-62181070

e-mail:chngpxygpfx@vip.sina.com; 修改稿专用邮箱:gp2008@vip.sina.com 网址:http://www.gpxygpfx.com